DEVELOPMENT OF A PREDICTIVE SPATIAL DISTRIBUTION MODEL FOR ERIODERMA PEDICELLATUM (BOREAL FELT LICHEN) FOR THE ISLAND OF NEWFOUNDLAND

RANDOLPH SKINNER

DEVELOPMENT OF A PREDICTIVE SPATIAL DISTRIBUTION MODEL FOR *ERIODERMA PEDICELLATUM* (BOREAL FELT LICHEN) FOR THE ISLAND OF NEWFOUNDLAND

by

© Randolph Skinner

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the

requirements for the degree of

Master of Science, Environmental Science

Faculty of Science

Memorial University of Newfoundland

April 2011

St. John's Newfoundland

Abstract

The worldwide population of *Erioderma pedicellatum* is currently listed as critical by the IUCN, with over 95% of the current population residing on the island of Newfoundland. Surveys *of E. pedicellatum* habitats and populations have primarily been opportunistic, rather than systematic in nature. Boreal felt lichen is listed as a species of special concern and vulnerable under COSEWIC and the Newfoundland and Labrador provincial listings respectively. By using a Geographic Information System (GIS) and compiled occurrence data and pseudo-absence data, I developed the first systematic predictive spatial distribution model for *E. pedicellatum* on the island of Newfoundland. A suite of 19 models using 4 different parameters were developed; the model with distance from coastline and aspect was the best candidate. Testing with reserve data and using a confusion matrix showed that the model displayed low model sensitivity (i.e., a low ability to predict false presence), but high model specificity (a strong ability to predict true absence). The final predictive model can assist future COSEWIC status assessments and provincial conservation management decisions that require information on probable species distribution.

Keywords: *Erioderma pedicellatum*, boreal felt lichen, predictive habitat model, presence absence model, GAM, Newfoundland

Acknowledgments

I would like to gratefully acknowledge the following people, groups and organizations for their kind input, information, advice and support that made my research possible. Thanks to my supervisors, Dr. Yolanda Wiersma and Dr. Mano Krishnapillai for their help, guidance and counsel during my research. I would also like to thank Bill Clarke, Dr. John McCarthy and Claudia Hanel for their expert opinions, insights and encouragement throughout my research. Thanks to Carl Marks and Joan Luther, for their technical expertise and advice regarding GIS-related issues. Thanks to Christine Doucet for initiating my interest in graduate school and for being a committee member. Thanks to Dr. Luise Hermanutz, Dr. Michele Piercey-Normore for their role as thesis examiners. Thanks and acknowledgements to the Newfoundland and Labrador Departments of Environment and Conservation and Natural Resources, the Canadian Forest Services, Memorial University of Newfoundland and Labrador and the Landscape Spatial Ecology Analysis lab (LESA) for providing me with equipment, software logistics and office space during my research. Funding for this work was provided by Memorial University's School of Graduate Studies, the Newfoundland and Labrador Department of the Environment, a GEOIDE grant to Dr. Wiersma, and student support funding to Dr. Krishnapillai through the Grenfell Campus. Infrastructure to complete the research was provided through Canadian Foundation for Innovation Grant to Dr. Wiersma and through in-kind support from the Wildlife Division and the Grenfell Campus.

Table of Contents

List of Tables

List of Figures

1 Introduction

In conservation biology and landscape ecology, trying to determine the spatial and population extent of rare species (in particular, rare species that avoid detection, or are difficult to detect in their natural habitat) is a challenge. By their nature, rare species offer us little natural life history as research may be scarce due to the difficulty of finding, identifying and maintaining contact with the species long enough to conduct viable studies (Pearson, et al., 2007). When the population of a rare species is in rapid decline, the challenges tend to increase. On a landscape level, studying the apparent decline of rare species with only a brief dossier on its natural life history requires the researcher to look at fundamental, basic environmental mechanisms that may determine limiting factors of the species' spatial distribution. Such is the challenge in determining the spatial distribution of the rare cyanolichen, *Erioderma pedicellatum (Hue) P.M. Jørg* (referred to as *E. pedicellatum*); known commonly as boreal felt lichen.

To develop a statistically valid means of predicting present and future spatial distributions for a given species, researchers and conservations managers can rely on a predictive model of spatial distribution. A predictive model is used to determine present and future trends in a variety of statistical situations, such as the development of anti-spam email systems, or future marketing trends (Kaplan and Urwitz, 1979, Bratko, et al., 2006). In ecology, a predictive model aids in determining present and future trends in species distribution, population dynamics or habitat suitability over a given landscape (Urban, 2000). A predictive model can therefore aid in conservation management decisions and in research and survey endeavours, and pinpoint potential population issues that need to be addressed (e.g., anthropogenic land development, population survey

planning) (Guisan and Thuiller, 2005). A predictive model consists of a number of predictors or variables that may influence the outcome of the model end-result in a statistically significant way (Guisan and Zimmerman, 2000). By developing the model on the basis of these predictors, the model can then be validated or refuted by comparing predicted values with real-life survey data. Finally, a predictive model is dynamic: as new predictor data become available, or as new predictors are tested, the model can change and, hopefully, become further refined (Guisan and Zimmerman, 2000). Ecological predictive models can use a combination of life history and environmental data. Most predictive models will include presence data, absence data, and one or more environmental or life history variables. Life history and environmental data can consist of biotic and abiotic factors that are measured when and where presence or absence data are taken, such as the degree of surface moisture, species migration patterns, size of territory, types of plants and animals in proximity, types of preferred substrates, temperature, and elevation (Franklin, 1995).

Most data sets on species distributions contain information on presence (occurrence) and some environmental variables. Ideally, one should include absence data (i.e., when a surveyor has not found the species at a given survey site) in the development of spatial distribution models. Absence data are used with presence data to give a clearer statistical indication of the habitats a species does and does not prefer. A model based upon known preferred and non-preferred habitats can then be used to look at a suite of similar, un-surveyed habitats found across a given landscape and assign probabilities that the species will be found there. However, most population surveys do not keep detailed records of absence data (Hirzel, et al., 2002). This may be because of survey time

constraints, lack of randomized survey techniques (opportunistic sampling), or the lack of immediate use of such data (Hirzel, et al., 2002). However, it is possible to develop predictive habitat models with presence-only data, using specific statistical models designed for this purpose (Elith, et al., 2006). Alternatively, 'pseudo absences', or randomly generated absence data, rather than surveyed absence data, can be used in binary models, especially in circumstances where species are extremely rare, or rare (Engler, et al., 2004). Pseudo-absence data have been successfully used for predictive habitat modeling for both mobile and sedentary species such as rare bat and fern species respectively (Sattler, et al., 2007; Zaniewski, et al., 2002).

After more than 30 years of study and opportunistic surveys on the island of Newfoundland, little is known of the life history of the boreal felt lichen. *E. pedicellatum*'s globally rare status makes it a good candidate for a predictive habitat model. The species is considered rare because it is difficult to detect. Because of this, it has been challenging to determine the underlying biology and ecology of the species that would give researchers a better idea of possible distribution (Mass and Yetman, 2002. Additionally, there has been no concerted effort until recently to record absence data. The objective of this thesis was to use presence and pseudo-absence data and four environmental factors that are hypothesized to limit boreal felt lichen distribution (topography, surface moisture, distance from coastline and desired habitat substrate) to develop a predictive spatial distribution model for boreal felt lichen on the island of Newfoundland. A final spatial distribution model could help conservation managers to determine possible desirable habitat area for future survey work. A predictive distribution map can also aid in assessing critical habitat areas or identifying areas for further research

or where more careful survey work for *E. pedicellatum* is required in advance of any proposed human development in those areas.

1.1 *E. pedicellatum* Biology and Ecology

A symbiotic relationship between a fungi substrate and an algae/cyanbacteria outer layer, a lichen represents one of the oldest forms of multicellular terrestrial life (Seaward, 1977). Lichens are a prolific form of composite organisms that can be found on all continents, and have a wide range of forms and highly diverse array of fungi-algal/bacterial associations (Seaward, 1977). The symbiotic relationship is comprised of the fungal species (mycobiont), which provides anchorage to a substrate, or phorophyte; while the algal species (photobiont) provides nourishment via photosynthesis (Seaward, 1977). Cyanolichens are poikliohydritic, enabling a fungus to adapt to extremes in dryness (Brodo, et al., 2001). When surface moisture is available, both the mycobiont and photobiont will produce metabolites that enable each constituent species to begin metabolism. Because it does not extract any water from the surrounding environment except for what falls directly on each individual lichen, biomass growth is dependent on surface and near-surface (atmosphere) moisture, and thus biomass growth only occurs during wet periods of the year.

E. pedicellatum is an epiphytic cyanolichen found in coastal boreal forests with a historic amphi-Atlantic distribution (Keeping and Hanel, 2006). Originally discovered in New Brunswick in the early 1900s, it has since become extirpated from New Brunswick (Keeping and Hanel, 2006). A detailed survey of Nova Scotia has been ongoing for several years, and to date, about 100 individual lichens have been found (Cameron and

Neily, 2008), although recently a single tree with over 50 individuals was discovered on Cape Breton Island (Ayers 2010). In northeastern Europe, it was believed the *E. pedicellatum* had a range throughout Scandinavia (Schiedigger, 2003). It is believed to be extinct throughout the Ferro Scandinavian range (Schiedigger, 2003). The exception to this global loss of diversity is on the Island of Newfoundland, where surveys have discovered in excess of ten thousand individual lichens, with more discoveries each survey year (Keeping and Hanel, 2006). The North American population of *E. pedicellatum* has been subdivided between the maritime populations, which now only consists of Nova Scotia, but was historically found in New Brunswick, and the boreal population on the island of Newfoundland (Keeping and Hanel, 2006). *E. pedicellatum* is comprised of the mycobiant *Erioderma* and the photobiont *Scytonema*. Separately, these mycobiant and photobiont species do not exist in the natural environment, and are only found together when *E. pedicellatum* is formed (Schiedegger, 2003). On the North American side of the Atlantic, *E. pedicellatum* has been found predominantly on balsam fir, *Abies balsamea* (Keeping and Hanel, 2006). In isolated cases, *E. pedicellatum* has been found on yellow birch, *Betula alleghaniensis*, and black spruce, *Picea mariana* (Schiedegger, 2003). In Europe, *E. pedicellatum* has been found on Norway spruce, *Picea abies*. On both sides of the Atlantic, *E. pedicellatum* has been found predominantly in wet, coastal forest climes (Schiedegger, 2003).

E. pedicellatum is leaf-like in appearance (Schiedegger, 2003). This leaf-like appendage or thallus is slate grey to blue-green in coloration with upward-curled grey edges. This coloration is dependent upon the amount of surface water retention (blue-green during wet periods seen in Figure 1). Mature thalli have distinctive fruiting bodies,

known as apothecia. These orange-brown fruiting bodies distinguish boreal felt lichen from other similar-looking lichens such as *Coccocarpia palmicola*.

Figure 1. Two photos of *E. pedicellatum* (courtesy of Newfoundland and Labrador Environment and Conservation) Note the white curled edges on the thallus (leaf-like regions) of both photos, as well as the reddish aphothecia prominent on the right thalli. *E. pedicellatum* on the right is in the dark green coloration (characteristic of wet periods). *E. pedicellatum* on the left is in a lighter grey phase (characteristic of dry periods), which tends to render the apothecia dark brown to black in color.

The lifecycle for an epiphytic lichen can be quite complex, with several growth and decay periods (referred to as necrotic stages) that can see the lichen advance and retreat in overall size, or lose its central portion altogether and form a ring-like pattern on a tree (Keeping and Hanel, 2006). This long and varied lifecycle may present the lichen with several challenges. In order for *E. pedicellatum* to sustain viable growth for a complete lifecycle, it must form and attach to a tree substrate that will last in excess of 30 years (Keeping and Hanel, 2006). Schiedegger (2003) notes that balsam fir seems to be the preferred tree substrate (both on the trunk and in the branches) of boreal felt lichen. It is believed that there may be a connection to the relative smoothness of a young balsam fir, which continues onwards for more than 30 years, and into middle and later aged trees. As the tree becomes venerable in age, the bark becomes more flaky and rugged, and it seems

that *E. pedicellatum* do not colonize well on older balsam fir trees. Keeping and Hanel (2006) suggest there is evidence that *E. pedicellatum* prefer not only balsam fir found within a forest stand, but may also prefer balsam fir trees near the edges of bogs and fens. These individual trees tend to grow much more slowly, and although they are smaller than balsam fir in a representative stand of trees, the slow growth permits the bark to stay smoother for longer periods of time, and the wet microclimate is believed to be conducive to *E. pedicellatum* habitat viability.

1.2 Brief Synopsis of the Boreal Population of *E. pedicellatum*

The first, informal survey to identify boreal felt lichen in Newfoundland was conducted in 1959 (Keeping and Hanel, 2006). Follow-up surveys on the Island of Newfoundland were conducted in the 1980s by T. Ahti and Wolfgang Maas (Ahti, 1983). Confirmed identifications on the Northern Peninsula, as well as on the Burgeo highway, Bay d'Espoir area, and the Avalon Peninsula were noted at the time, but no follow-up surveys were conducted for the next 15 years. At the time, spatial distribution of *E. pedicellatum* was relatively unknown throughout all of the North America ranges, although a rapid decline was noted by observers throughout all of the original surveyed regions on both sides of the Atlantic (Keeping and Hanel, 2006). By 1998, there had been a dramatic decline in overall species populations and distributions worldwide. With a more than 90% reduction throughout all of its historical maritime population ranges, and all-but-one known individual lichen in the Ferro-Scandinavian region of Europe in 2003, *E. pedicellatum* was included in the International Union for Conservation of Nature (IUCN) Lichen Red List as critically endangered worldwide. To date, boreal felt lichen is

listed as endangered in the maritime population range. On the Island of Newfoundland, *E. pedicellatum* has been listed as a species of Special Concern under the COSEWIC (Committee on the Status of Endangered Wildlife in Canada), and Vulnerable under the Province of Newfoundland and Labrador's *Endangered Species Act* (Maass and Yetman, 2002; Keeping and Hanel, 2006). Therefore, special care needs to be undertaken when any human development occurs in areas that are believed to harbour populations of *E. pedicellatum*, under the provincial guidelines of the Species at Risk Act and Species at Risk Policy (Government of Newfoundland and Labrador, 2010). For *E. pedicellatum*, no development can occur in current habitat, and any forest stand that is deemed a possible candidate habitat for *E. pedicellatum*, must be surveyed before any human development can occur (Hanel and Keeping, 2006).

1.3 Current Status of *E. pedicellatum*

Keeping and Hanel (2006) state that concerted survey efforts for boreal felt lichen began on the Island of Newfoundland in 1998. The provincial Department of Natural Resources (at that time, under the auspices of the Department of Forest Resources and Agrifoods), adopted a landscape management approach on the premise of retaining prime habitat at and near known boreal felt lichen populations. Under the landscape management approach, it was hoped adjacent forest stands that continued natural growth and cycling processes would act as both a buffer and new potential habitat for currently known *E. pedicellatum* populations.

During the surveys from 1998-2008, two distinct hyper-populated regions were mapped in detail (Figure 2). The Avalon population is centered in the heart of the Avalon

Peninsula within the Lockyer's Waters region, while another is centered near the Jipujijkuei Kuespem Provincial Park in the Bay d'Espoir region. The combined populations within these regions comprise nearly 96% of the known individual thalli currently accounted in Newfoundland (24% and 72% of the entire known population respectively). In total, more than six thousand thalli have been surveyed and an estimated 10 thousand or more individual thalli are estimated to live within the two hyper-populated regions (Keeping and Hanel, 2006).

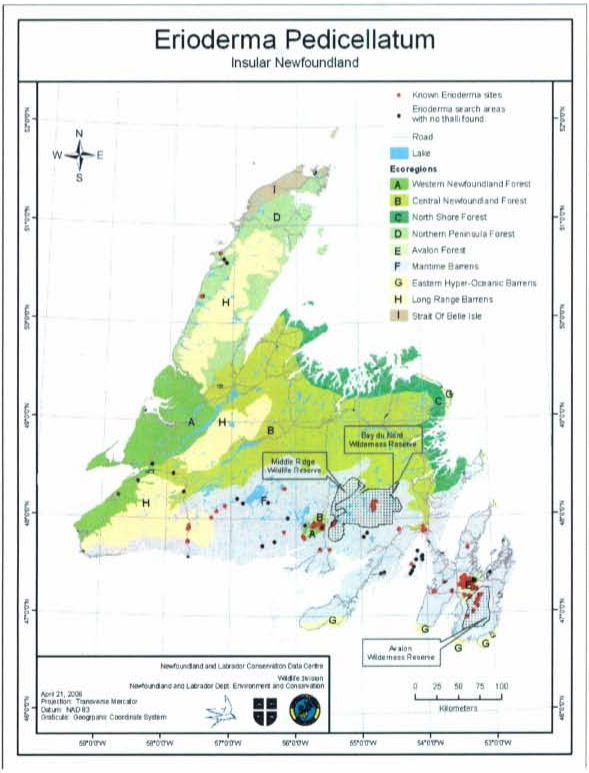

Figure 2. A map of the island of Newfoundland outlining *E. pedicellatum* sites in red (source: Keeping and Hanel, 2006, government of Newfoundland and Labrador, 2006).

In 1998, the Department of Forest Resources and Agrifoods adopted recommendations to create buffer zones around known *E. pedicellatum* populations and around potential adjacent habitat in the hope that this would ensure present and future *E. pedicellatum* populations could persist, and that potential colonization habitat was readily available (Bill Clarke, pers. comm.). The Forestry Division relied heavily on opportunistic, non-randomized survey patterns to establish such buffer zones (Ahti, 1983). Initial survey efforts focused primarily in, and adjacent to, previously known *E. pedicellatum* populations that were near designated cutting zones as a precaution to determine *E. pedicellatum* presence in and near those zones. Further surveys were done in regions that provincial authorities thought may have potential as prime habitat for *E. pedicellatum* (primarily the presence of balsam fir), and were slated as potential future cutting zones (Bill Clarke, pers. comm.). The provincial Department of Environment and Conservation's Wildlife Division has surveyed existing *E. pedicellatum* sites, and has furthered the census and surveying data over the last nine years with ever-expanding search areas, particularly near the provincial parks and protected areas that are near the current hyper-populated regions (Lockyer's Waters and Bay d'Espoir). The Conne River Miawpukek First Nation obtained federal funding via Environment Canada's Species at Risk Critical Habitat Protection Program for further monitoring and surveying efforts near the adjacent Jipujijkuei Kuespem Provincial Park, and has been working in tandem with the provincial Wildlife Division in their surveying efforts over the last five years. Most recently, the provincial Department of Environment and Conservation conducted a systematic survey of lichen species on the Avalon Peninsula, which was led by Dr. John

McCarthy (McCarthy, pers. comm.). To date, this is the only systematic field survey of lichens completed in the province (McCarthy, pers. comm.).

Both Clarke (pers. comm.) and Keeping and Hanel (2006) outline that surveys are opportunistic at best, as candidate sites are generally chosen based on the assumed habitat preferred by the boreal felt lichen (i.e., stands of balsam fir, moist microclimates, age and size of trees). Furthermore, surveys have been conducted primarily from roadside and trail access routes near the hyper-populated regions. The reasoning behind the lack of systematic surveys for *E. pedicellatum* has been one of logistics and funding. Newfoundland is largely unpopulated, and accessing areas even a few kilometers from roads and trails can be time consuming and difficult. The terrain is rocky and hilly, with many gulches, ravines and rivers crisscrossing the landscape. Surveys by helicopter are costly. Coastline access is difficult, with fog and wind in the spring, summer and fall, and ice during the winter.

In 2006, the Government of Newfoundland and Labrador initiated its 5 Year (2006 – 2011) Management Plan for the Boreal Felt Lichen (*Erioderma pedicellatum*) in Newfoundland and Labrador. One of the key aspects of the management plan's objectives is continuing research of the ecology and biology of the boreal felt lichen (Keeping and Hanel, 2006). It is hoped that by defining the ecological and biological life history of boreal felt lichen, definable management objectives can be developed and implemented. The management plan's implementation stage outlines the need to develop a predictive habitat model. Such a model can potentially assist in management issues and decisions about boreal felt lichen conservation at an island-wide extent (Keeping and Hanel, 2006). With a reliable predictive model, spatial distribution of boreal felt lichen

may be better understood, and aid in provincial government decisions on forest and land development, as well as forest harvesting issues that may impact known boreal felt lichen populations or those habitats which the model suggests may have a high probability of being suitable for boreal felt lichen (Keeping and Hanel, 2006).

1.4 Predictive Modeling of *Erioderma pedicellatum*

The amount of accumulated data from surveys of *E. pedicellatum* in Newfoundland is quite extensive (Keeping and Hanel, 2006). However, except for a minute amount of data from the 2009 Avalon lichen survey, there are no null survey points, or absence data (John McCarthy, pers. comm.). To develop a predictive model, it would be best to include both presence and absence data.

I reviewed the data from expert surveyors and researchers on the prevalent environmental factors that were prominent in the majority of boreal felt lichen finds. I narrowed the environmental factors to four primary predictors to include in the model: aspect, distance from coastline, host tree species and relative ground/surface moisture content. These four factors are hypothesized to have a significant impact on boreal felt lichen habitat, and two of these factors (tree substrate and distance from coastline) have been included in a heuristic model in the province of Nova Scotia as shown (Cameron and Neily, 2008). I decided not to use sphagnum moss wetlands in my model, as the data for such wetlands in the FRI and other GIS sources for the island are very poor and not suitable for the study (Table 1).

Table 1. a comparison model parameters of the Nova Scotia heuristic model and the systematic model for boreal felt lichen on the island of Newfoundland.

Nova Scotia Model	*Newfoundland Model*
Balsam fir within stands	Balsam fir within stands
Less than 30 km from coast	Less than 30 km from coast
within 80 meters of sphagnum moss wetlands	High moisture levels present
	Aspect of hills/terrain

On a stand level, across the island of Newfoundland, I hypothesize that the habitat suitability for *Erioderma pedicellatum* will be significantly correlated with (a) distance from the ocean (indicative of maritime climates), (b) the presence/absence of balsam fir, and (c) topography (specifically aspect). As one moves further inland from the coastline (ocean), the degree and regularity of moist maritime climate will decrease. I predict that habitat suitability will decrease as one moves further into the hinterland as shown in Figure 3 (Baldwin and Bradfield, 2005).

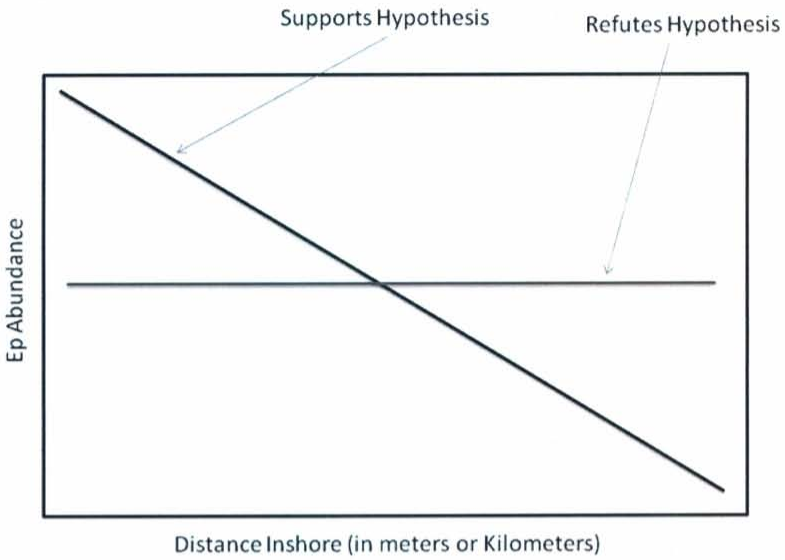

Figure 3. The hypothesized relationship between *E. pedicellatum* abundance as a function of distance inland from oceanic coast.

Studies and surveys indicate balsam fir seems to be the niche tree species where *E. pedicellatum* is consistently detected on the island of Newfoundland (Keeping and Hanel, 2006). Therefore, the presence of balsam fir may be an indicator of *E. pedicellatum* habitat suitability and the basis of my second prediction (Lang, et al., 1980).

E. pedicellatum habitat suitability is predicted to be higher within valleys, and lower near hills and peaks (Bill Clarke, pers. comm.). Moisture retention and collection at the bottom of valleys and on protected slopes may increase potential habitat suitability (Rolstad, 2001). Testing the *E. pedicellatum* habitat suitability against aspect, and TCI (Topographical Convergence Index) or TRMI (a TCI variant known as Topographical Relative Moisture Index) may determine a trend in abundance relative to these factors. (Figures 4 and 5). TCI measures the slope and aspect from a digital elevation model,

calculates the speed and collection areas in lower terrain (where water runoff collects in lower elevation, flatter terrain), and provides an index of low-to-high moisture collection at the surface. The TRMI model uses the TCI data, but then further refines the TCI parameters it to provide relative moisture or wetness levels near the surface beyond the collection areas (Parker, 1982).

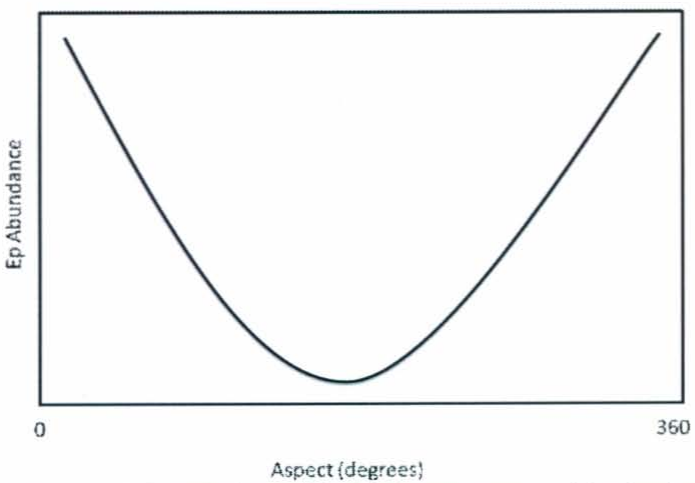

Figure 4. The hypothesized relationship between aspect and the abundance of *E. pedicellatum*.

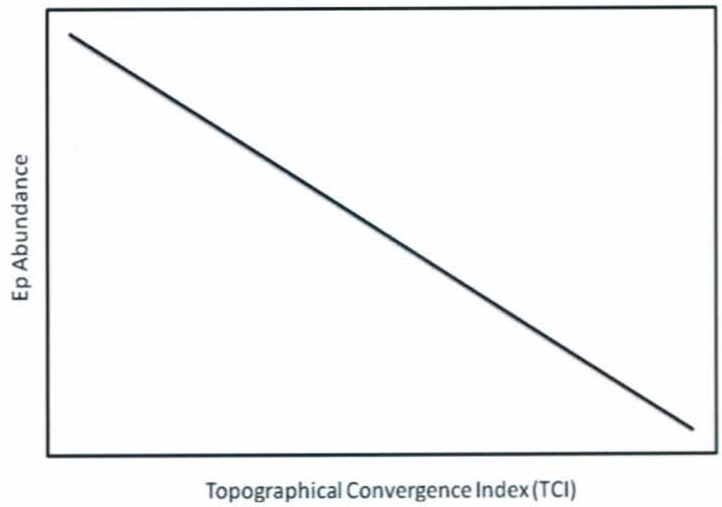

Figure 5. The hypothesized relationship between TCI (or TRMI) and the abundance of *E. pedicellatum*.

In addition to atmospheric moisture, moisture near the ground (surface moisture) may also be an important predictor (Bill Clarke, pers. comm.). Some of the variation in surface moisture may be captured in measurements of slope/aspect, but topographic convergence and relative moisture indices may supply more detail. Because lichens appear to thrive in damp environments (Brodo, et al., 2001), I predict that suitable *E. pedicellatum* habitat will more likely be those areas with high TCI and TRMI values.

I used these four environmental predictors to develop a predictive habitat model for *E. pedicellatum* for the island of Newfoundland. If the model has high predictive power, it may be a useful tool determining habitat suitability and estimated population distribution of the boreal felt lichen across the as-yet un-surveyed portions of the island.

2 Methods

To develop a statistical population distribution model for *E. pedicellatum* I filtered data compiled by a host of observers and surveyors across the province. Then I used the data to develop and validate the statistical model. Finally, I used the statistical model and implemented it into a Geographic Information System (GIS) to display the predictive surface visually on a map. This process was carried out in five stages. First, I assessed the *E. pedicellatum* data, and filtered data points based on the criterion of minimum grid size (see details below). Second, I developed randomly created absence data points (known as pseudo absence data) to offset the deficit of real absence data. I compiled GIS layers for each of the four predictors, and sampled all presence and pseudo absence data points for these using ArcGIS 9.3 (ESRI, Redlands, CA), and then divided the entire dataset into "training data" for development of the statistical model, and "testing data" for model validation (Fielding and Bell, 1997). Third, using the training data, I built a series of competing models, based on the hypotheses outlined above, and statistically determined which parameters best predicted the data (model building) (Anderson, 2008). Fourth, I tested the best model with the testing data (model validation) and created a confusion matrix to determine overall model error fitness (Fielding and Bell, 1997). Finally, I developed a predictive surface for probability of *E. pedicellatum* on an island-wide basis. Details of each of these steps are outlined below.

2.1 *Erioderma pedicellatum* data

I used data compiled over eight years by the Department of Environment and Conservation-Wildlife Division and the Department of Forest Resources and Agrifoods

to test my hypotheses and build predictive models. The data on *E. pedicellatum* occurrence are sampled as either point or area counts (Figure 6). Point counts include the actual GPS coordinates and the tree species substrate for each individual lichen thallus for each tree. Area counts note groupings of lichen and substrates for a given area surveyed for a given amount of time. Neither the spatial extent of the survey area, nor the time spent surveying are included in area counts (Bill Clarke, pers. comm., Claudia Hanel, pers. comm.). Thus I was not able to estimate the extent of the survey area from these data. I included all point counts in the final data set because each one represented a specific point on the map (i.e., accurate to well within 78.9 metres). To be consistent, the predictive model is based upon the smallest divisible GIS cell used in my study. The DEM map provided by the Canadian Forest Service had the highest resolution (smallest grain) of 78.9 meters. Because I was unable to estimate the time or area surveyed in the area counts, I could not reliably position the area survey data within a particular 78.9 m grid cell.

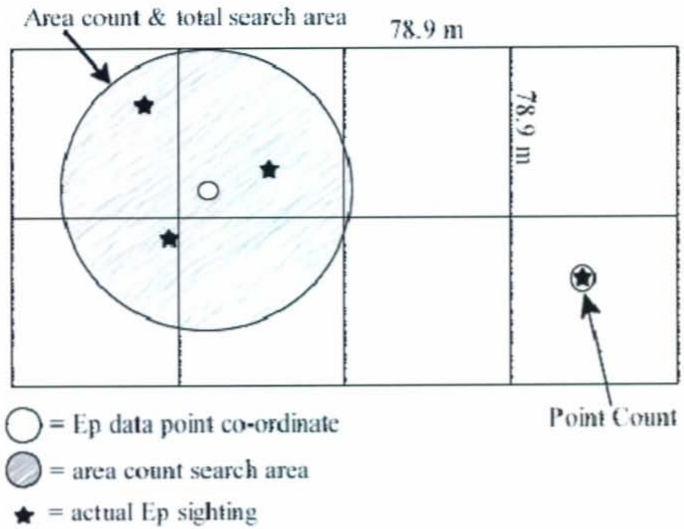

Figure 6. Area counts occupy the actual *E. pedicellatum* sighting space (represented by the star, which is merely centimeters in size), as well as the survey region (grey area). An area count can occupy a spatial region that is larger than the 78.9 meter grid criterion, but as a datum, it shows only one indivisible data point (represented by the hollow circle). Point counts occupy the actual space of the *E. pedicellatum* thallus (mere centimeters) as well as being represented in the data as a data point that is likewise indivisible for the purposes of my study.

Thus, the area counts were completely excluded from the analysis. I filtered the remaining data to include only one observation per 78.9 m grid cell, leaving a total of 667 points (mostly clustered in the Bay d'Espoir, and Lockyer's Waters regions of the province) Of these, I randomly chose ~10% ($n = 67$) of the points to reserve as testing data, leaving 600 points that represent *E. pedicellatum* occurrence in the training data set, which was used for statistical model building (Fielding and Bell, 1997).

2.2. Pseudo absence data

I used pseudo absence data in my study as a proxy for real absence data. Because collection of occurrence data for *E. pedicellatum* on the island of Newfoundland has been somewhat opportunistic and focused on areas of known high density, systematic collection of absence data (i.e., notation of where surveys did not find *E. pedicellatum* when following the same search protocols) was not done until 2009 (John McCarthy, pers. comm.). Pseudo absence data has been used in other studies of rare, or difficult to find species. Thus, in order to develop a presence-absence model, I generated pseudo absences to offset the lack of absence data (Wisz and Guisan, 2009). I created random points across the entire forested part of the island of Newfoundland in ArcGIS 9.3 (ESRI, Redlands, CA, 2008) using Hawth's Tools Random Sampling Toolset; Beyer, 2004). I constrained these random points to forest cover that included balsam fir in the stand (see details on Forest Resource Inventory data below). Initial inspection of the presence and pseudo absence data showed that all presence data were within 20 km of the coast (Figure 7).

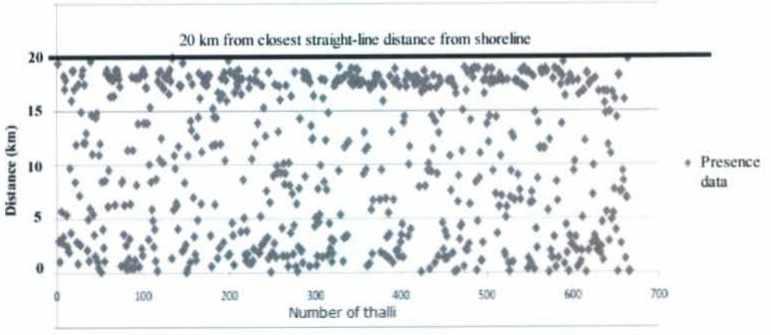

Figure 7. Scatter plot of the presence data in relation to distance from coastline. Note the 20 km ceiling for the furthest filtered data point from the coastline.

This may be a real biological constraint (i.e., *E. pedicellatum* does not occur beyond 20 km from the coast), or it may be an artifact of the opportunistic sampling locations, all of which were in close proximity to the coastline (Claudia Hanel, pers. comm.). Nonetheless, to avoid statistical overfitting, I decided to build my statistical models for *E. pedicellatum* locations within a 20 km zone inland from the coast, around the entire island (Figure 8).

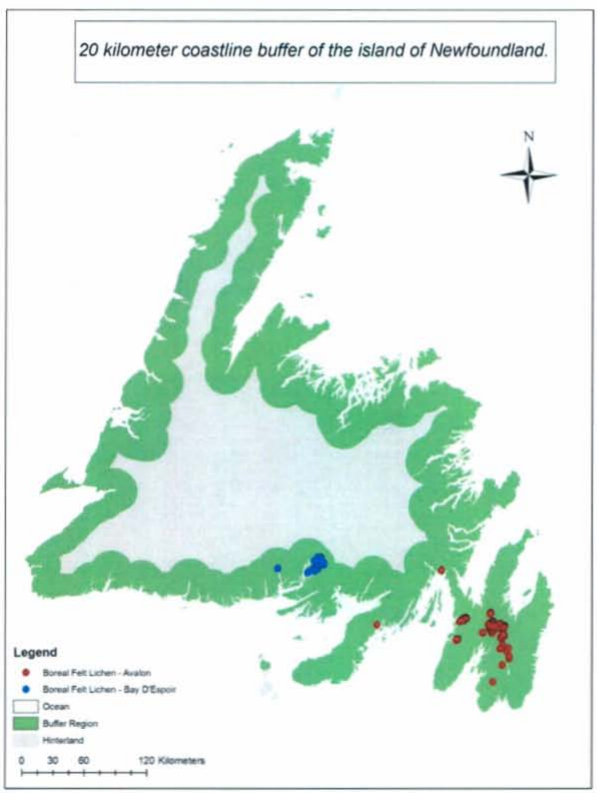

Figure 8. Island of Newfoundland, with a 20 km buffer (green shaded region). Note the extent in the Bay d'Espoir (center map) and Avalon (right hand portion of map) that contain most of the existing surveyed population of *E. pedicellatum* found to date.

Within all forest stands containing balsam fir and located within 20 km of the coast, I generated approximately twice as many pseudo absence points to real presence, since a 2:1 ratio of pseudo absences to presence data has been suggested to increase the statistical power, and to increase the likelihood the data would give a statistically significant representation of true absence data (Wisz and Guisan, 2009). Due to the nature of the random point generation across the island, ArcGIS 9.3 did not develop a perfect 2:1 ratio of pseudo absences. My final count was n=1180, and it was deemed close enough to the 2:1 ratio to proceed. I randomly selected 10% (n = 118) of these points to reserve for testing data, leaving 1062 pseudo absence points in the training data set, for a total presence + pseudo absence data set of n = 1662.

2.3 GIS layers

2.3.1 Forest Resource Inventory

I used the provincial Forest Resource Inventory (FRI) as a primary GIS layer to determine forest stand structure (provided by the Government of Newfoundland and Labrador, Department of Natural Resources). FRI data provide a spatial data set for forest stands based on regular surveys and are digitally compiled for use in GIS (Gillis, 2001). The focus of the FRI is on quantifying forest stand productivity for timber harvest, but these types of data have been used in spatial ecology research in the past both in this province and in other parts of the world (Moores, et al., 1996; Gillis, 2001). The database classifies stands based on dominant species, stand composition (i.e., non-dominant species), average tree age and height, and percentage of canopy closure. *E. pedicellatum*

are known to predominantly occur on balsam fir (*Abies balsamea*) but have also been found occasionally on yellow birch (*Betula alleghaniensis*) and black spruce (*Picea mariana*) (Keeping and Hanel, 2006). However, when the presence data set was used to sample the FRI data with an overlay analysis in the GIS, all of the data points occurred in stands that were dominated by or that contained Balsam fir. Thus, I constrained pseudo absence data (which represent "available" habitat) to those stands that contained Balsam fir.

2.3.2 Distance from coastline

I used the "Near" Tool in ArcGIS 9.3 Proximity Analysis toolbox to measure distance from each point to the nearest section of the coastline, using a coastline GIS layer of the island of Newfoundland.

2.3.3 TCI and TRMI Layers

To approximate near surface and surface moisture for the model input, I used proxy models developed to determine surface water retention and magnitude. The Topographic Convergence Index (TCI) and the Topographical Relative Moisture Index (TRMI) were developed by S. Wilds (Wilds and van Manen, 1995; Wolock and McCabe, 1995) as a way to determine surface moisture for the Appalachian Mountains, the US Midwest and the Rocky mountains for GIS-based gradient analysis and niche modeling (Wolock and McCabe, 1995). It has since been used worldwide and has been accepted as a robust proxy for surface moisture values when 'real' moisture data are unavailable or not reliable (Wolock and McCabe, 1995). It is a relatively straightforward modeling procedure that only requires a Digital Elevation Model (DEM; a raster file containing

elevation values in each cell) as the input. By using slope, aspect and steepness gradients from a DEM, the TCI and TRMI models can calculate relative topographic convergence and infer areas where surface moisture accumulates and drains away.

I built TCI and TRMI layers for the entire province in ArcInfo with Arc Macro Language (AML) scripts that were previously developed to calculate TCI and TRMI for the Appalachian mountain range. I calculated TRMI in whole number format ranging from 7 (lowest moisture value) to 60 (exceptionally high near surface moisture). For the TCI, the data ranged from 0 (high topographic convergence, and hence high surface moisture) to 14.7 (low topographic convergence and hence low surface moisture).

Wilds and van Manen's (1995) original models for TCI/TRMI were built for the Appalachians in the eastern United States. To determine whether the TCI and TRMI parameters in van Manen's original model were suited for the province's varied terrain, I conducted a sensitivity analysis on both models. Parameters from both models were changed by a magnitude of 5 (i.e. 5x larger than and 5x smaller than the actual value) and then re-run for the province. For both the TCI and TRMI, all possible parameters were tested for sensitivity. TCI parameters included flow area and flat surface determination. The TRMI parameters included a combined planiform and profile curvature, relative slope potential (RSP), slope and aspect. Only one parameter was changed at a time, and then each model was separately run in ArcInfo. If the expected TCI or TRMI values varied considerably from the original (non magnified parameter values), further investigation would have been required to determine the relative deviance of sensitivity for the suspected parameter.

Once I was confident that the TCI/TRMI models were robust to the permutations introduced in the sensitivity analysis, I sampled the TCI/TRMI values at each data point using the Extract Values to Points tool in ArcGIS 9.3 Spatial Analyst's Extraction toolbox. Figures 9 and 10 show TCI and TRMI respectively graphically on digital maps using ESRI ArcGIS 9.3.

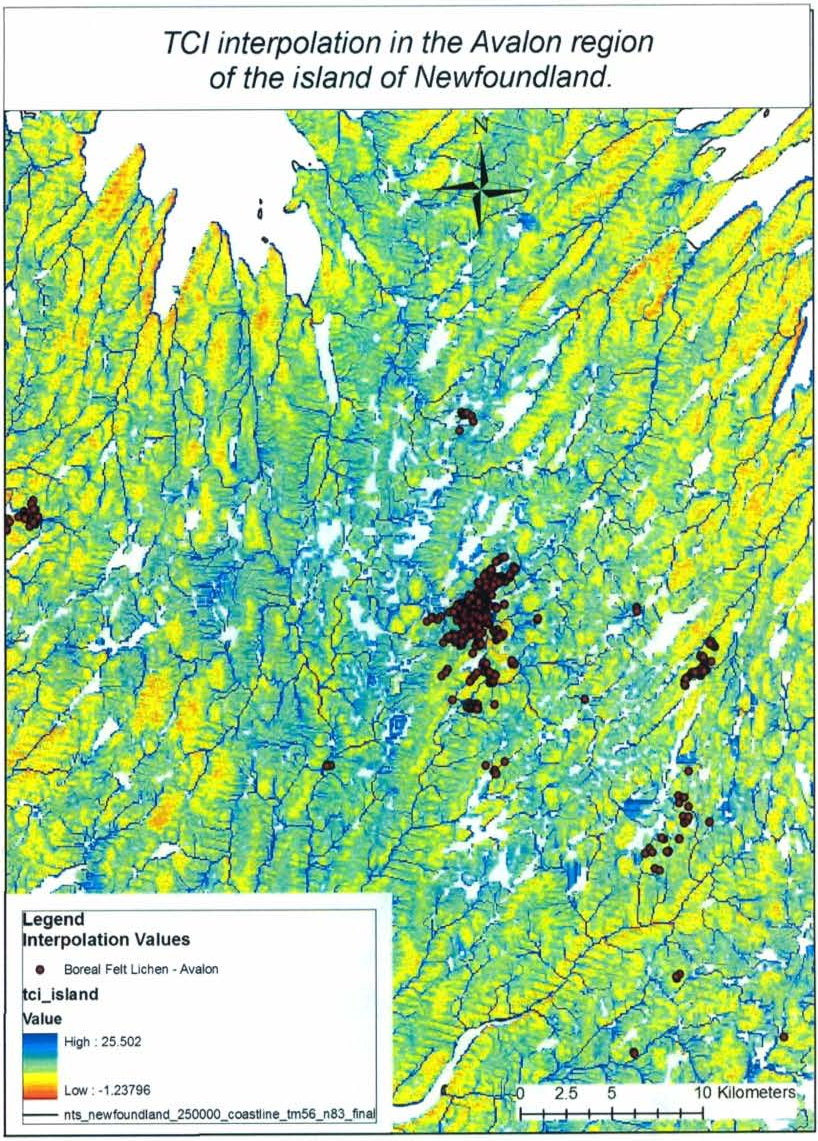

Figure 9. TCI values in the Avalon region of the island of Newfoundland. Blue denotes possible areas of greatest surface moisture convergence, while red denotes poor convergence.

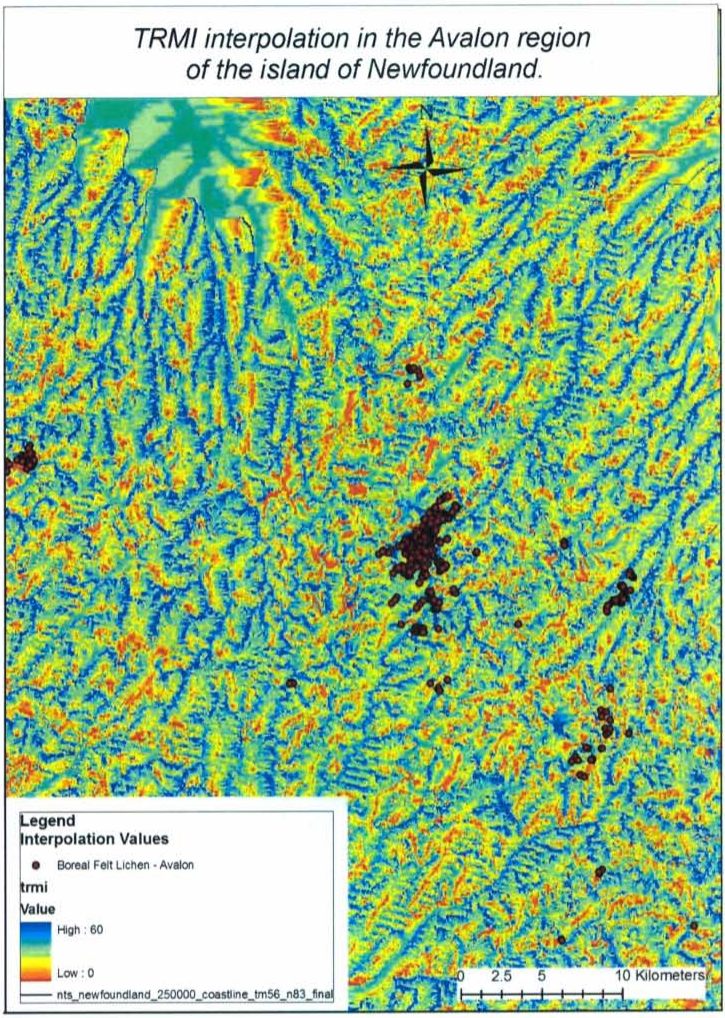

Figure 10. TRMI interpolation values in the center of the Avalon peninsula. Blue denotes wetter regions, while red denotes drier regions.

2.3.4 Aspect

I created one final GIS layer to represent aspect. I created the aspect layer from the DEM using the Aspect tool in ArcGIS 9.3 Spatial Analyst's Surface Analyst toolbox.

I extracted Aspect values at each data point using the Extract Values to Points tool in ArcGIS 9.3 Spatial Analyst's Extraction toolbox and then recalibrated values from a 0-360° to a scale of ±180° to give "northern" values the same order of magnitude (without this, aspects a few degrees east or west of north differ by a value of 360° (e.g., 5° east of north is 5° while 5° west of north is 375°. However, in terms of solar insolation (which is what aspect is a proxy of), these values should be equivalent. With recalibration, their magnitude is ±5°). Figure 11 shows aspect output via ESRI ArcGIS 9.3.

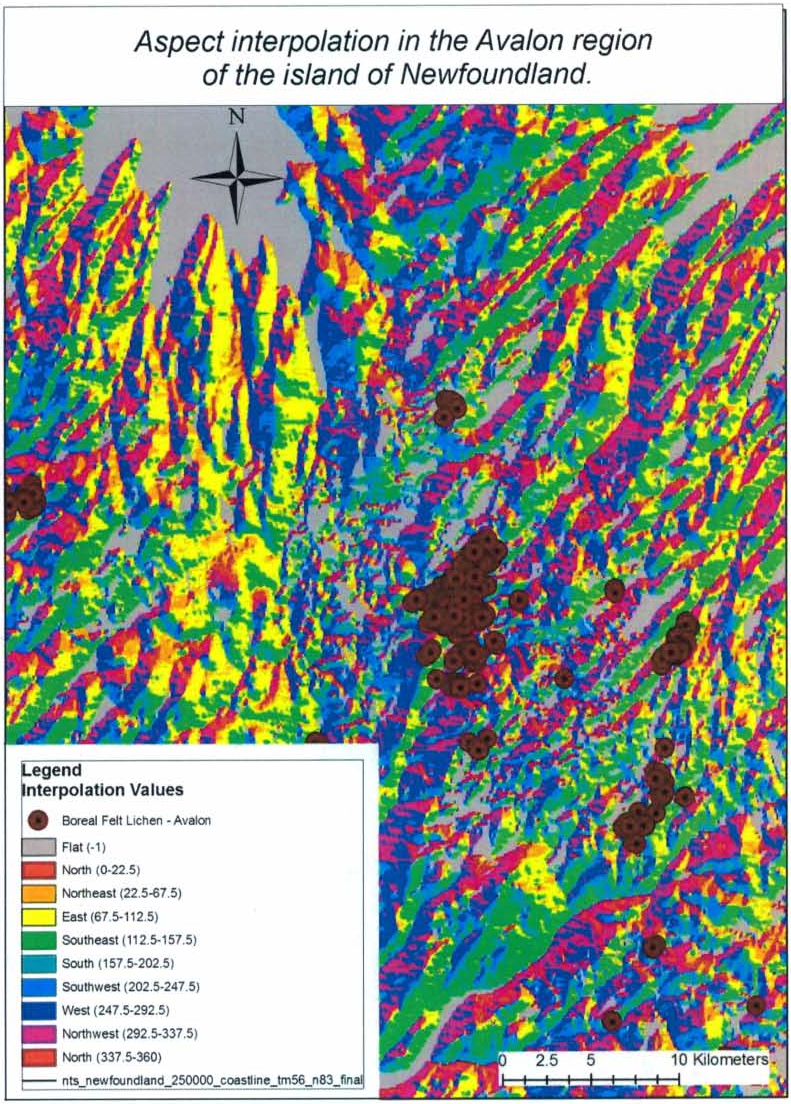

Figure 11. Topographical aspect in the center of the Avalon region of the island of Newfoundland. Each color represents a change in aspect. The boreal felt lichen data points have been enlarged to provide easier identification.

2.4 Statistical Analysis

I carried out statistical analysis in two phases as suggested by Fielding and Bell (1997). In the first phase (model building), I subjected the training data to a series of statistical tests to determine (a) normality of the data (b) any possible transformation(s) of the data to provide better statistical normality (c) the implications of normal versus non-normal data and (d) determination of the best statistical model that could be used for the given data. In the second phase (model validation) I applied the best model from the first phase to the testing data to create a confusion matrix to determine the specificity and the sensitivity of the model. Sensitivity is the measurement of the amount of positive (presence) measurements taken that are actually correct, while specificity relates to the number of negatives (absences) that are correctly identified (Fielding and Bell, 1997).

2.4.1 Model Building

Data were compiled in Microsoft Excel, and then each model predictor was tested for normality. Of the three predictors chosen, moisture (both the TCI and TRMI data sets) showed Poisson distribution. The TCI and TRMI data sets were transformed using the square root of each datum. Aspect and Distance from Coastline, however, had bimodal distributions. Bimodal distributions are generally dealt with using non-normal (non-parametric) statistical analysis.

For model development, I opted to use the Generalized Additive Model (GAM), a non-parametric counterpart to the parametric General Linear Model and the Generalized Linear Model because of the non-normality of the data, particularly the bimodality of the

distance and aspect predictors. GAM is widely used in biological and ecological model development when data are not normal and data transformation is not practical (Guisan et al., 2002). Using a backfitting (retrofitting and replacing) algorithm, the Generalized Additive Model differs from the traditional Generalized Linear Model by replacing the weighted linear regression in the adjusted dependent variable by a weighted backfitting (one that is repetitively replaced) algorithm (Hastie and Tibshirani, 1986). This fitting iteratively smoothes partial residuals by separating the parametric portion of the fit from the nonparametric portion, fitting the parametric portion using weighted least squares within a Gauss-Seidel backfitting algorithm (Hastie and Tibshirani, 1986).

What makes GAM very useful is the ability to pick and choose which predictors use the non-parametric smoother and which ones do not. This allows for a more robust model, as non-parametric (smoothed) data is dealt with differently than the parametric (non-smoothed) data, all within the same model. I used the default parameters for the GAM model as outlined in the R Statistical package (versions 2.9.0 to 2.12.0) with the MGCV library (Hastie, 2010). It is recommended that unless warnings are generated in the R program, or if the model fitting fails to converge, the default parameters be used (Hastie, 2010).

2.4.2 Model selection

Originally, I had chosen five predictors to develop my model: species substrate (S), distance from coastline (D), Topographical Convergence Index (T1), Topographical Relative Moisture Index (T2), Aspect (A). These were the physical parameters from my original hypothesis.

The full model, therefore, was:

$$O = S + D + (T1 \text{ or } T2) + A$$

where O represents presence-absence occurrences, with a 1 (presence) or 0 (absence). My filtered data showed that 100% of all occurrences happened only on Balsam Fir. Thus, the parameter Species (S) was omitted from the model, as its assumed inclusion is inevitable (i.e., zero explanatory value). Thus, my new base model is O = D + A + (T1 or T2). From this initial model, 18 other *a priori* models were developed (Anderson, 2008) as outlined in Table 2. These models were then subjected to a series of statistical significance tests based on the Generalized Additive Model (GAM).

I statistically analyzed the model selection using the R statistical software package (versions 2.9.0 to 2.12.0, 2009-2010 respectively, R Project for Statistical Computing). Each of the 19 models was analyzed using the Generalized Additive Model library (with the MGCV library). For each model, the predictors were individually evaluated on their p-values, as well as ranked based on their respective adjusted R^2 values, deviance explained and the UBRE scores (which is a modified Akaike's Information Criterion value that measures model fitness). I set my final p-value significance test at p=0.05. Those models that had all significant predictors were then evaluated on their R^2 adjusted values and their deviance explained values. The adjusted R^2 value gives us insight on how much of the model is statistically explaining the real-life biological and ecological factors, while the deviance explained is a rough guide that contrasts the model with the 'full' model, or a model with all parameters fully fitted. Both a high R^2 value and deviance explained value would fare well in the final model

selection; the higher R^2 and deviance explained values the better statistical fit the model had.

Models that had all significant predictors, and had relatively high R^2 and deviance explained values were then further evaluated using a modified Chi-Squared ANOVA test using the GAM (mgcv) library package (anova.gam). Models that had two significant predictors, with a third predictor that was nearly significant (i.e., near, but just over the 0.05 *p*-value significance), and with relatively high deviance explained and R^2 values were included, just to be sure that all significant values (*p*-values, R^2 values and deviance explained values) were taken into account. The ANOVA chi square test determines the most significant model from the suite of models (see Table 2 below).

Table 2. Potential *E. pedecillatum* models for the island of Newfoundland and their respective statistical analysis including when applicable, chi square, *p*-values, z-values R^2 adjusted and deviance explained. T1 and T2 denote TCI and TRMI moisture models. The s denotes the smoothing function used in the Generalized Additive Model.

Model	Equation	Predictors	Chi Sq	P-Value	Z-Value	Deviance Explained	R2 Adjusted
1	O = sD + sA + T1	sD	380.06	2.00E-16	na		
		sA	16.79	0.36	na		
		T1	na	0.092	1.68	32.30	0.37
2	O = sD + T1	sD	385.40	2.00E-16	na		
		T1	na	0.070	1.81	31.30	0.36
3	O = sA + sD	sA	380.85	2.00E-16	na		
		sD	16.93	0.033	na	32.20	0.37
4	O = sA + T1	sA	22.10	0.0065	na		
		T1	na	0.20	1.29	1.23	0.010
5	O = sD	sD	386.10	2.00E-16	na	31.20	0.37
6	O = sA	sA	20.15	0.017	na	1.13	0.001
7	O = T1	T1	na	0.74	0.33	0.01	-0.00050
8	O = sD + sA + T2	sD	380.91	2.00E-16	na		
		sA	19.93	0.012	na		
		T2	na	0.072	1.80	32.3	0.37
9	O = sD + T2	sD	386.80	2.00E-16	na		
		T2	na	0.25	1.15	31.2	0.37
10	O = sA + T2	sA	20.22	0.012	na		
		T2	na	0.53	-0.30	1.14	0.0095
11	O = T2	T2	na	0.76	-0.30	0.01	-0.0005
12	O = sD + sA + sT1	sD	330.30	2.00E-16	na		
		sA	18.31	0.021	na		
		sT1	29.06	0.0061	na	33.70	0.39
13	O = sD + sT1	sD	332.66	2.00E-16	na		
		sT1	28.68	0.00071	na	32.70	0.3770
14	O = sA + sT1	sA	17.12	0.030	na		
		sT1	111.76	2.00E-16	na	6.84	0.078
15	O = sT1	sT1	115.50	2.00E-16	na	5.87	0.070
16	O = sD + sA + sT2	sD	381.10	2.00E-16	na		
		sA	20.05	0.012	na		
		sT2	3.37	0.066	na	32.40	0.37
17	O = sD + sT2	sD	386.82	2.00E-16	na		
		sT2	1.30	0.26	na	31.2	0.37
18	O = sA + sT2	sA	20.20	0.012	na		
		sT2	0.37	0.55	na	1.14	0.0095
19	O = sT2	sT2	0.10	0.75	na	0.01	-0.00050

2.4.3 Model validation

I validated the final model using the testing (reserved) data (Fielding and Bell, 1997). In R, the reserve data was used in the model using the GAM (MGCV) library and the predict.gam command. Figure 12 outlines the process I used. The prediction command took each reserve data point, input it into the selected model, and output a value that was statistically weighted to determine how each value 'fit' or was indicative of the model's parameters. Each data point was scored on how well it fit to the ideal output for the chosen model. Further, the prediction data were used to generate a probability of occurrence based on the exponential link with a logistical transformation to weight each value between 0 and 1. Finally, values were then assigned a 0 or 1 if they were below or above 0.5 value respectively. These values were then assigned to a confusion matrix (Fielding and Bell, 1997) to assess model fit.

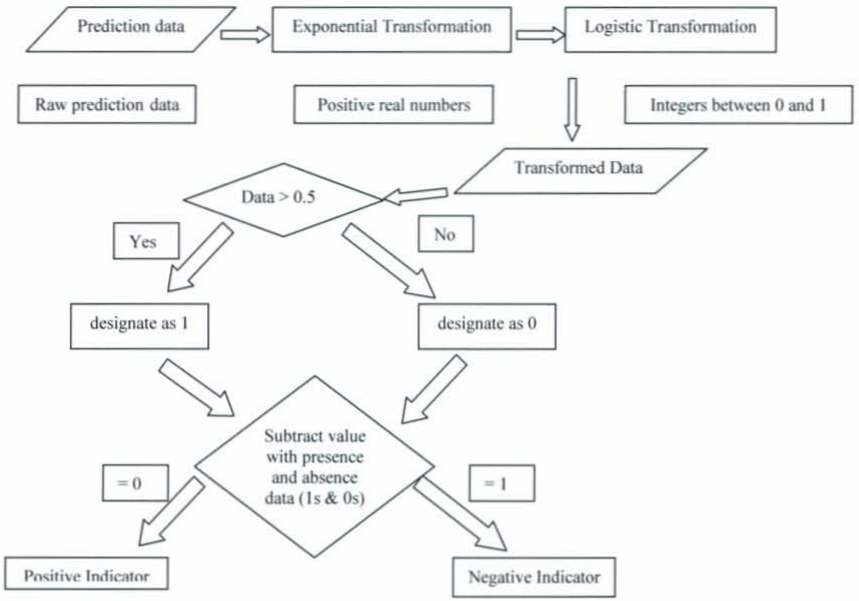

Figure 12 Flow chart outlining the transformation of prediction data generated into R, into the final *data set* denoting positive or negative indicators of the presence and absence data.

2.4.5 Predictive Surface

Once the confusion matrix for the test data was complete, I developed a predictive surface. In ArcGIS 9.3, I created a lattice of data points separated by 1 km using Hawth's Tools, Sampling Tools and Generate Sampling Points command. Hawth's tools restricted the lattice to cover the province from the coastline to 20 kilometers inland. These data points were then sampled for their respective aspect and distance data, the data were then extracted from each datapoint in ArcGIS 9.3 and converted to a spreadsheet in Microsoft Excel. The prediction.gam command was used in R to produce the predictive values for

each data point. The same transformation treatments (prediction value to exponential to logistic) were done for the data as was completed for the training data as outlined in Figure 12. The logistic values were then exported back to the ArcGIS 9.3 layer in the same data set as the 1 km grid data points. Using the Spatial Analyst toolbox, I then created an inverse distance weighted interpolation based upon the newly imported logistic values. The final output was a predictive surface of the island of Newfoundland based upon the best model.

3 Results

3.1 Sensitivity Analysis

None of the parameters tested using a step change of ±5X exaggeration of the parameter values displayed any significant change in output values. Thus both TCI and TRMI passed the sensitivity analysis (Table 3).

Table 3. Percentage of change in each parameter's final value when the variable was changed to a given magnitude from the pre-set norm (in this case, the normal variable value is given a magnitude zero).

	Magnitude						
	-10	-5	-2	0	2	5	10
Parameter							
Flow Accumulation(TCI)	-5	0	0	0	0	0	5
Flat Surface Area(TCI)	0	0	0	0	0	0	0
Flow Accumulation(TRMI)	0	0	0	0	0	0	0
Focal Mean(TRMI)	0	0	0	0	0	0	0
Top Threshold(TRMI)	0	0	0	0	0	0	0

3.2 Final Model Selection

Figure 13 outlines the final model selection process using a flow chart. Based on the ANOVA chi square test the final model selected was model 3: $O = sD + sA$, where O

is occurrence, s represents the smooth (GAM) function, the D is distance from coastline and A is the aspect of the slope. The next closest model (model 12 which was O = sD + sA + sTCI) also had all the predictors significant, and had the best R^2 values and deviance explained values (Table 3). However, in the Chi-Squared ANOVA test, O = sD + sA proved significant over O = sD + sA + sTCI. This confirms the principle of parsimony, as the model with fewer parameters (the chosen 2-parameter model) as more squared bias, but inherently has less uncertainty than the rejected 3-parameter model (Anderson, 2008).

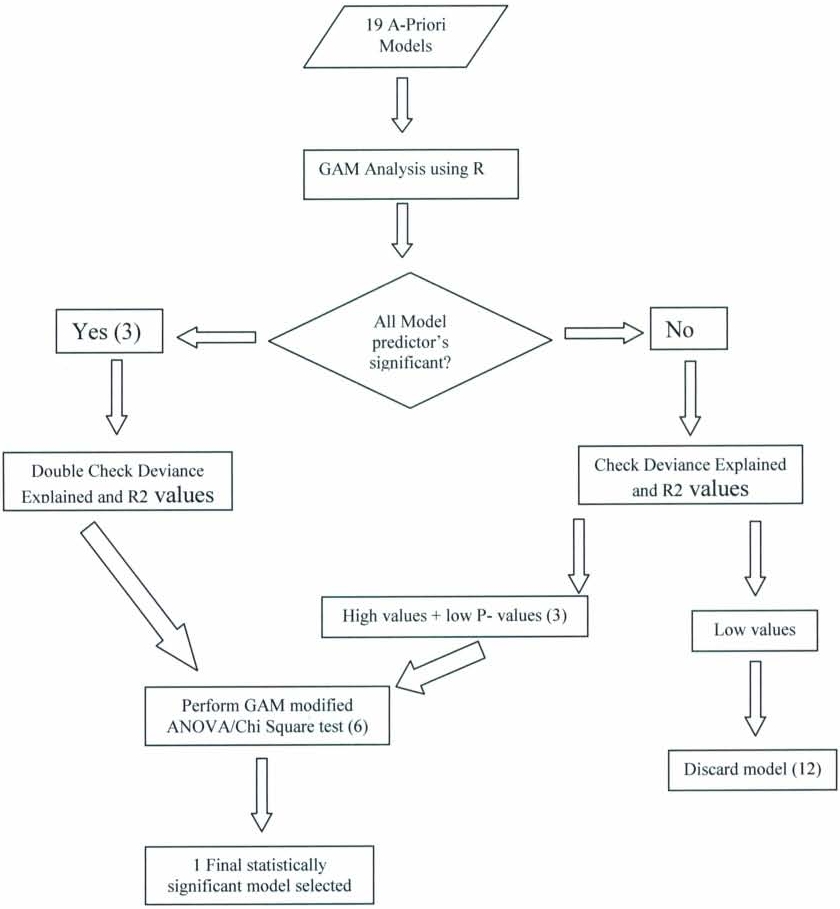

Figure 13. Flow chart process in choosing the final model in creating the predictive surface from a suite of 19 potential models. Numbers in parentheses indicate the number of models during a given process.

The shape of the relationship between boreal felt lichen occurrence and the predictor values of distance from coastline, aspect, TCI and TRMI are shown in Figures 14-17.

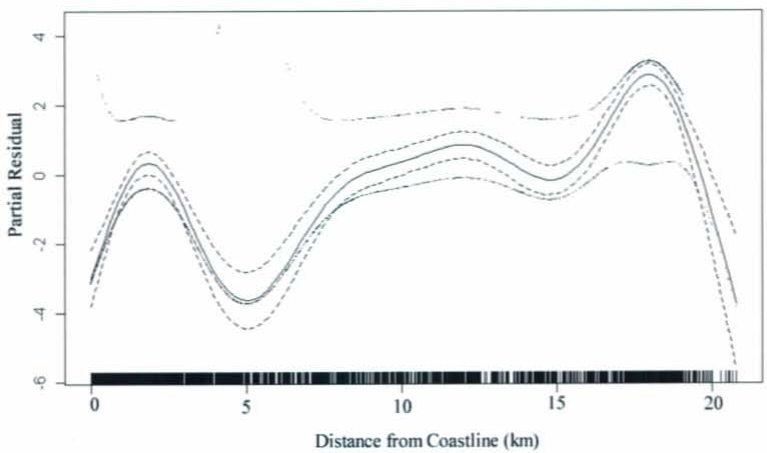

Figure 14. Distance from coastline versus partial residuals on the y-axis, using the generalized additive model (GAM) the dotted lines indicate the confidence interval (95%), while the solid line represents the distance from coastline partial residual.

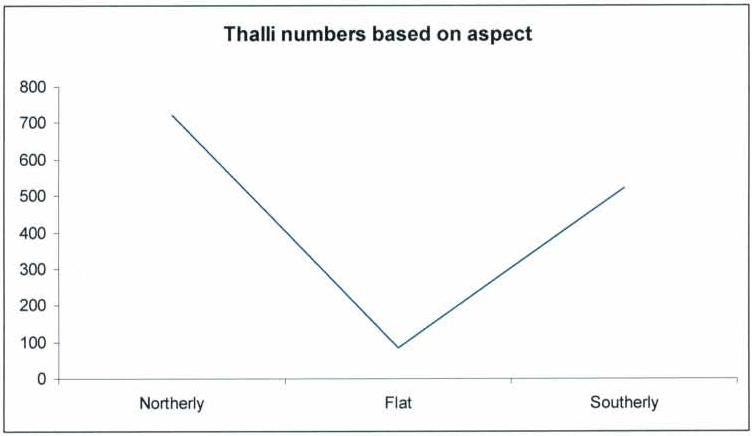

Figure 15. Boreal felt lichen occurrences (from filtered data) based on aspect. From my data set, more boreal felt lichen were found on the north facing rather than south facing slopes or on flat terrain.

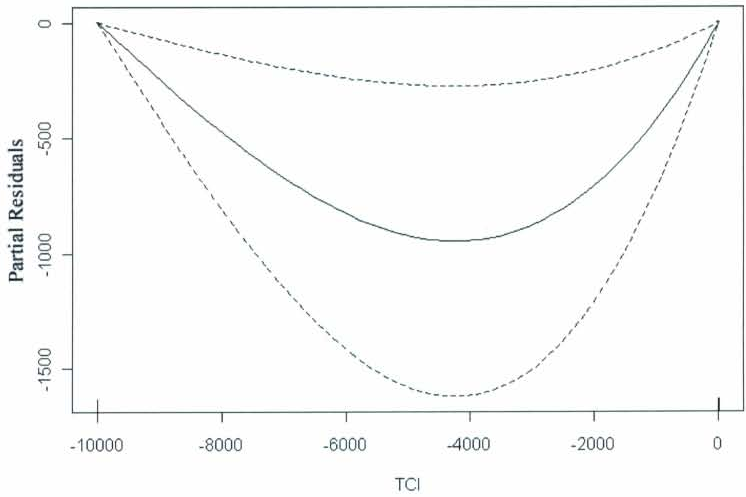

Figure 16. Actual TCI values with partial residuals plotted on the y-axis, using the generalized additive model. The solid lines are the partial residuals while the dashed lines represent the confidence interval (95%).

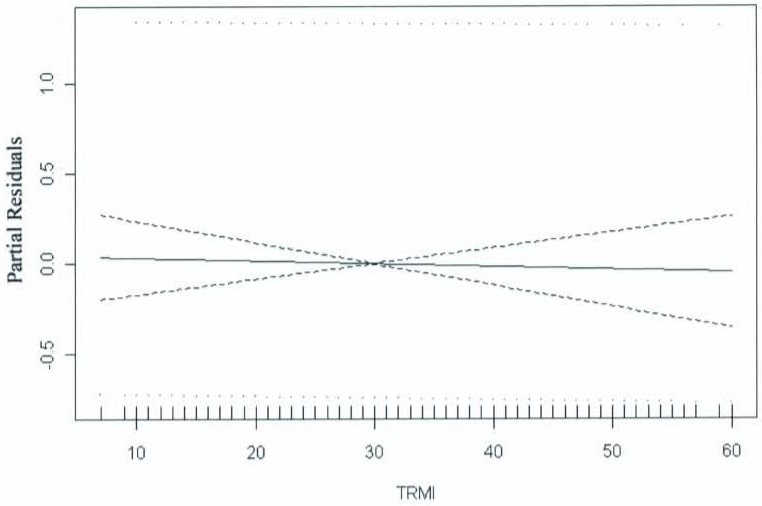

Figure 17. Actual TRMI values with partial residuals (solid lines) plotted on the y-axis and the 95% confidence intervals (dashed lines), using the generalized additive model (GAM)

3.3 Determining Model Fitness

Fielding and Bell (1997) explain that confusion matrices are used to determine the number of prediction errors in any given model. Prediction errors are misleading results for presence and absences and can be either false positives or false negatives. A confusion matrix not only provides a measure of false negatives and false positive values, but a suite of statistical analysis providing a better indication of the model strength.

Table 4. Confusion matrix outlining the predicted and actual positive and negative values, as well as key values and their respective percentages for Model 3.

		actual	
		+	-
predicted	+	10	33
	-	57	84
		67	117

	%
Prevalence	36.41
Overall Diagnostic Power	63.59
Correct Classification Rate	51.09
Sensitivity	14.93
Specificity	71.79
Kappa	9.24

Results of model validation are shown in Table 4. Sensitivity and specificity outlines the accuracy of the presence data and absence data respectively. Sensitivity is low, at 14.93%, which indicates that the model is not predicting the presences of *E. pedicellatum* without the risk of a type I error (Fielding and Bell, 1997). However, with a specificity of 71.79%, the model is giving a good percentage of correct predictions for the absence data. Overall diagnostic power is at 63.59%, which indicates the model lacks predictive accuracy. The correct classification rate is at 51.09 %, indicating that the model has only a 50% chance of correctly classifying predicted presence and absence. Kappa is very low

(9.24%) which indicates that the model has overall poor agreement with the criteria set out by Fielding and Bell (1997) for model fitness.

3.4 Surface

The predictive surface for the entire island is shown in Figure 18. Detailed maps for the prediction surface from each of the two areas of the province where intensive survey work for *E. pedicellatum* have occurred are shown in Figure 19 (Avalon) and Figure 20 (Bay d'Espoir). Figure 18 represents the predictive model surface as applied to the island based on the abiotic parameters, and does not restrict the suitable habitats to only those that overlap with Balsam fir stands, because at that resolution, the Balsam fir stands are indistinguishable. However, within the pixels coded as "highly probable habitat" in Figure 18, only those that overlap with stands of Balsam fir would be considered probable *E. pedicellatum* habitat.

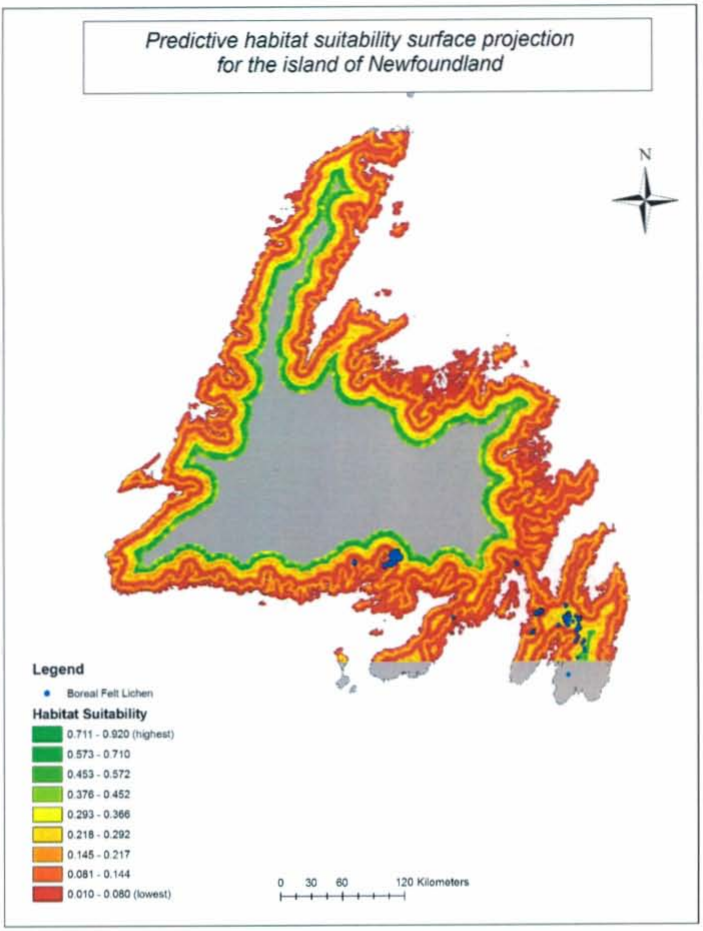

Figure 18. The final island-wide predictive surface interpolation. Dark red represents the lowest probability of suitable *E. pedicellatum* habitat, and bright green represents the most probable. Numeric values represent the statistical predictor values for the model, with 0.920 being the highest, and 0.010 being the lowest. The blue data points represent currently known *E. pedicellatum* populations that were used for this study. Grey areas denote regions not interpolated by ESRI ArcGIS9.3. The Balsam fir stands were not represented on this map, as the resolution is too low to be shown effectively. For maps that include the Balsam fir stands, see Figures 19 and 20.

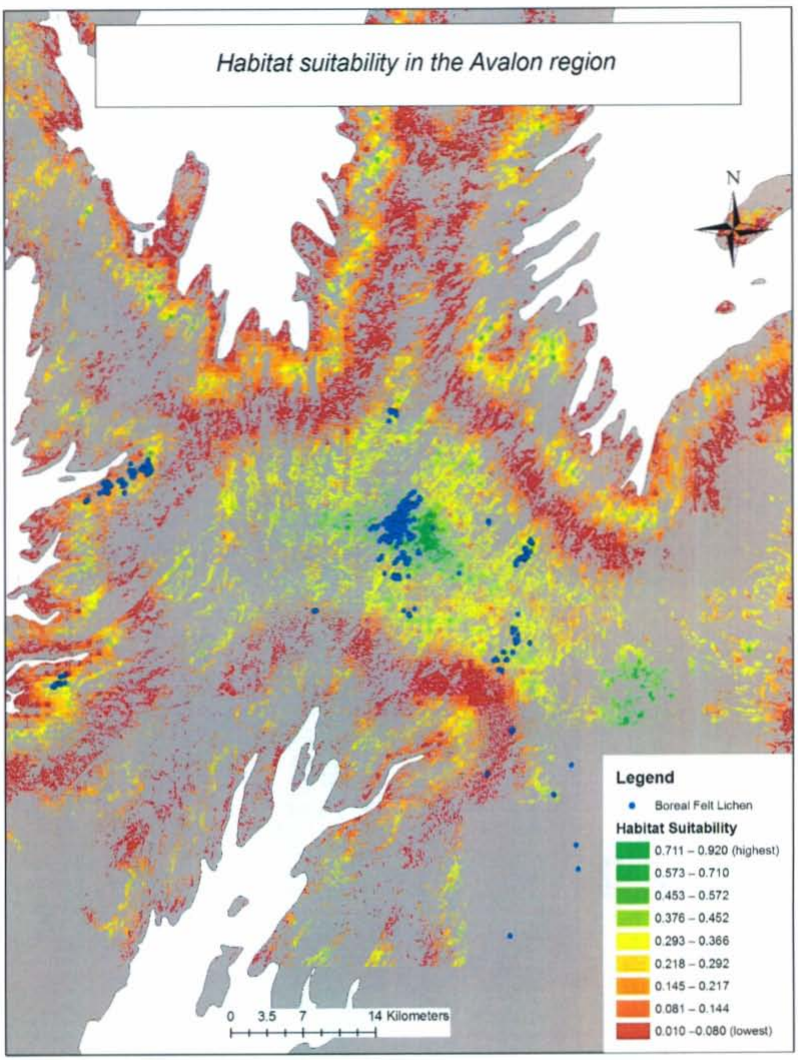

Figure 19. The final predictive surface interpolation, focusing on the Avalon region of the province shown only within balsam fir stands. The brighter green regions and higher numbers represent more suitable predicted habitat, while darker red regions and lower numbers show increasingly less suitable habitat. Grid cells are at 500 meters to due to computer memory limitations.

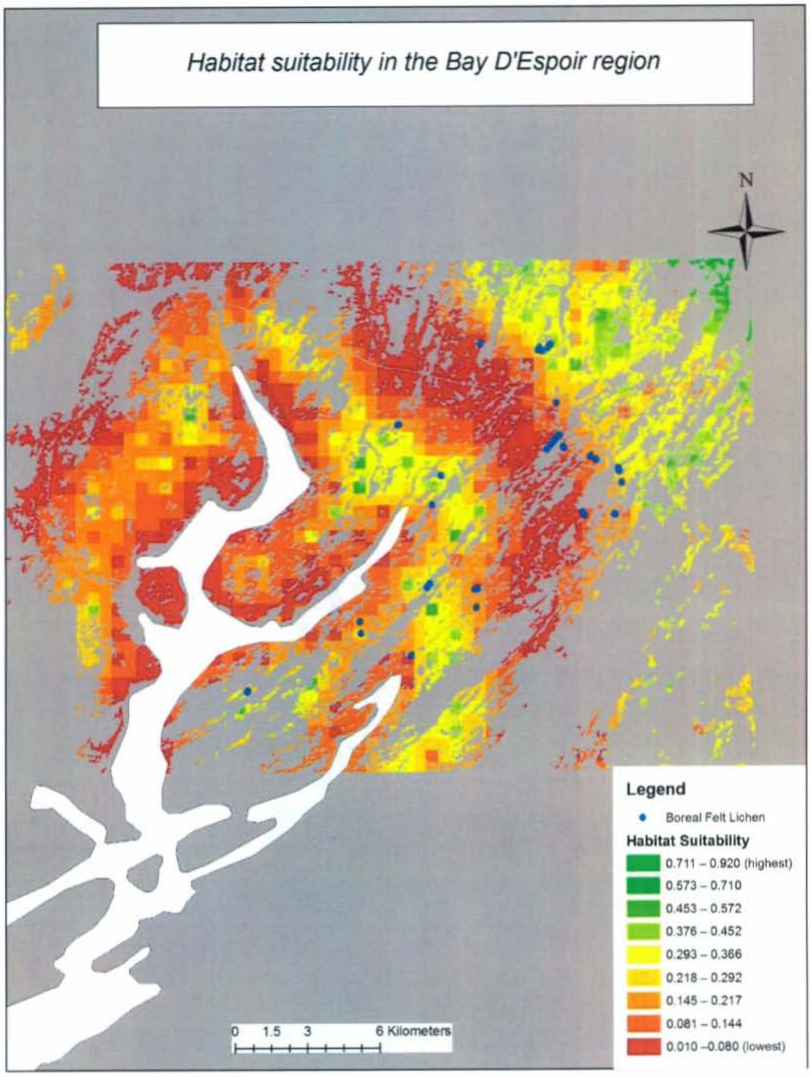

Figure 20. The final predictive surface interpolation, focusing on the Avalon region of the province shown only within balsam fir stands. The brighter green regions and higher numbers represent more suitable predicted habitat, while darker red regions and lower numbers show increasingly less suitable habitat. Grid cells are at 500 meters due to computer memory limitation.

4 Discussion

The final predictive model was the best of 18 possible *a priori* models I had selected at the beginning of my research. The model parameters for all candidate models were based on field research experience spanning more than a decade, and were picked because they were consistent both in the database and from expert opinion. Because this is an initial predictive model for *E. pedicellatum*, the initial parameters met two criteria: (1) they could be easily obtained from existing data sets, and (2) meet the requirements of both the minimum grain size (78.9 meters) and be translated to a GIS software computer program. The best of all candidate models, given the data, contained only the aspect and distance from coastline parameters.

For comparison, the *E. pedicellatum* prediction model developed in Nova Scotia used a heuristic, or 'rule of thumb' model development and selection (Cameron and Neily, 2008). A heuristic model is not statistically based and is aimed more at explaining a pattern through direct observational patterns, whereas a predictive model (such as the one developed here) examines underlying statistical significances within a given set of parameters in order to develop a final prediction model for a set of proven statistically significant parameters. The heuristic model developed for Nova Scotia used distance from coastline (less than 30 kilometers), and balsam fir within 80 meters of sphagnum wetlands as their primary biological and geographical predictors using a filtering query developed in a GIS. In this thesis, I investigated a suite of possible predictors statistically, and then applied the statistically best fitted model via GIS. In my analysis and in the Nova Scotia model distance from coastline was a strong predictor. I opted not to include distance from wetlands as a possible statistical predictor in my initial model

development, as the geospatial data for sphagnum moss wetlands in the FRI were not detailed enough province-wide.

Aspect was non-parametric, so I could not statistically determine which direction was most significant, but only that aspect in general was statistically significant. On a cursory level, aspect did show an approximate preference for sloped terrain over flat terrain, as well as northerly slopes over southerly slopes. Aspect was significant within the final model, but discussing relative direction can be problematic. The preference of *E. pedicellatum* habitat for sloped rather than flat terrain may be in part to limit the amount of direct sunlight reaching inside the tree canopy and to the boreal felt lichen communities therein (Gauslaa, et al., 2001). The loosely inferred preference of northern slopes over southern slopes may also be a result of the amount of ambient sunlight filtering through the forest canopy during the day (Campbell and Coxson, 2001, Hylander, 2005). Alternately, it may be a preference to greater near surface moisture retention from a northern aspect versus a southern aspect or it may simply be a case of substrate dependence, as north facing balsam fir trees will have a slower growth life cycle than the southern facing trees. Hanel (pers. comm.) indicated that the retardation of the balsam fir growth during its full lifecycle may give more opportunities for the microhabitat substrate to retain features that are conducive to a more complete boreal felt lichen life cycle. Further research would have to be done to confirm the relationship between rate of balsam fir growth and *E. pedicellatum* substrate preferences.

The distance from coastline predictor showed a solid, bimodal indication that near the coastline (approximately 1 to 3 kilometers) and at the edge of the coastline 20 kilometer buffer (approximately 16 to 18 kilometers) was the preferred habitat of boreal

felt lichen in the two test regions. This may be due to several biotic and abiotic gradient factors. First, the population density of the preferred substrate (balsam fir) also fits the pattern shown in the boreal felt lichen population densities and distance from coastline. Sudden or abrupt changes in the Daman types from the coastline to the interior may explain this phenomenon (Meades and Moores, 1989). Furthermore, exposure to direct coastal climatic conditions during all four seasons (salt water, ice and higher winds) may be detrimental to boreal felt lichen habitat suitability directly at the coastline (Werth, et al., 2005). Balsam fir that grow closest to the shoreline may exhibit stunted growth, or be exposed to the oceaninc elements year-round.

While statistically, the model containing aspect and distance from coastline was the most significant model, the confusion matrix of agreement and fit showed that the final model was not particularly robust. It had a sensitivity of only 14.9% and specificity of 71.8% and a Kappa reading of only 9.2%. The very low Kappa reading can be explained by the disparity of data used for the presence and pseudo absence. Fielding and Bell (1997) indicate that when one category of data outnumbers another by a significant amount, the Kappa reading becomes less reliable. In my study, pseudo absence data outnumbered presence data by a ratio of 2:1, but was necessary to ensure better overall model fitness (Wisz and Guisan, 2009). This would definitely undermine the final Kappa value and therefore, render it of little worth for the final model fitness analysis. However, with the underlying poor value for sensitivity and overall diagnostic power (63.9%) and correct classification rate (51.9%) are good indicators that the model needs some future work to refine, add or remove parameters to improve overall predictive fitness.

The model had low sensitivity and the model fitness was generally low. Two of the four possible predictors were statistically significant, and became the basis of the final model. However, as with most rare and hard to find species, knowledge about the biology and ecology can be either lacking or in its infancy. A lack of solid biological or ecological background may lead to poor model fitness (Anderson, 2008). Therefore, my model, while relatively poor for determining where *E. pedicellatum* can be found, is relatively good at determining where it is likely to be absent.

There are several reasons why my model had such poor overall predictive power (but was relatively good at predicting absences). At present, the majority of *E. pedicellatum* data are restricted to two primary hyper-densely populated regions: the Bay d'Espoir enclave on the south coast and the Lockyer's Waters region of the Avalon Peninsula. One of my key predictors for the model was distance from coastline, which showed a clear bimodal distribution both very near to the coastline (less than 2 kilometers) and at the edge of a 20 kilometer range. At first glance, this looks to be a solid relationship between coastline distance and population density. However, further analysis can detect these trends are caused by geographical and biogeographical nuances of Newfoundland's shape and floral population distribution. First, geographically, the island of Newfoundland is comprised of several large, yet narrowly shaped peninsulas. One of the largest is the Avalon Peninsula, where the Lockyer's Waters population resides. Looking at the data from the Avalon Peninsula, and determining the distance from coastline, we can see that no place on the peninsula is greater than 20 kilometers from the coastline, with the Lockyer's waters region being the furthest inland (i.e. near the 20 km region). This may explain the bimodality with the Avalon data. The Avalon

data also comprised the bulk (70%) of the data used in the study, therefore a definite bias towards displaying distance trends would come out of the overall data set. The Bay d'Espoir region, while not constrained geographically as the Avalon Peninsula, does have biogeographical considerations. From the Damman charts and the Forestry Resource Index, balsam fir stands quickly thin out to bogs and fens, and eventually to black spruce dominated regions beyond the 20 kilometer range from the coastline in the region (Meades and Moores, 1989). Because balsam fir is the primary substrate for *E. pedicellatum*, the relatively quick change in tree species composition could mark the hinterland limit of *E. pedicellatum's* range. This would further bias the data towards a 20 kilometer limit that is limited by physical geography (Avalon) and host substrate (Bay d'Espoir) constraints.

A further challenge in model development was defining moisture at or near the surface at a province-wide extent. Field researchers have widely believed that air-borne humidity and moisture levels could be one of the most important population factors for *E. pedicellatum*. The challenge was finding some form of moisture recording or derivation at a high enough resolution to correspond to *E. pedicellatum* distribution. Using historical data from weather stations, such as the federal Environment Canada's weather data was deemed inadequate for the study, as it would have been too coarse. Using the Topographical Convergence Index and Topographical Relative Moisture Index (TCI and TRMI) was deemed a satisfactory proxy for near surface fog and humidity levels. While the models proved successful in emulating surface moisture levels (and have been used successfully in other research endeavours), they were not significant predictors for *E. pedicellatum*.

There may be other statistically significant predictor(s) of *E. pedicellatum*'s population distribution which were not included in my study. The predictors I chose were based upon expert advice from years of field work researchers working with current and past *E. pedicellatum* populations in the province of Newfoundland and Labrador. There is the possibility that other environmental or habitat predictors have been overlooked because they are difficult to collect, or data are lacking, or they are too complex to articulate as a single predictor in a GIS environment (Table 1). Several potential predictors such as temperature, tree age, adjacency to bogs and fens, and the inclusion of other lichens in the nearby ecological community might be likely predictors of *E. pedicellatum* occurrence. These, however, were deemed not suitable for my current study due to time constraints, lack of real data or data that did not conform to the scale of my study, and the lack of a field research component of my thesis. For example, even though temperature on an island-wide scale would be useful for the study, to predict if seasonally high or extreme low temperatures possibly limit population ranges, the lack of accurate monthly or seasonal temperatures at the resolution I required for my study meant that I could not use that parameter.

The amount of presence data garnered by provincial government and First Nation's researchers and surveyors over the last decade is quite impressive. However, the bulk of the data (~80%) was based on site counts, rather than point counts. As I outlined in my methods section, all of the site counts had to be removed from my data set because (1) uncertainty of the exact search area per site count, and (2) no recorded data on time of search effort or approximate area searched. Because my minimum grain was 78.9 meters, this meant that many data points (in excess of 80%) were filtered out of my study. These

filtered data points may have had an impact on statistical power. What impact the increased sample size would have had to the final model's selection or overall statistical fitness and power could not be determined.

Finally, the lack of true absence data may have had an impact on modeling success. Pseudo absence data have been used when true absence data have been lacking in several case studies over the years. Pseudo absence data have also been used for rare and hard to find species. The issue is how statistically significant are the pseudo absence data for *E. pedicellatum*, when compared to actual absence data. For my study, I had less than 15 absence data records at hand, which did not comprise a significant portion of my overall pseudo absence data.

4.1 Limitations and Recommendations

For further research into the landscape ecology of *E. pedicellatum* I recommend at least five possible improvements to the model development, research techniques and other possible candidates for a more improved model. Most of the recommendations stem directly from the challenges and issues I have mentioned above, as solutions to these issues. Some, however, are recommendations to research new avenues with regards to prediction modeling *E. pedicellatum* on the island of Newfoundland.

Geographic and bio geographic constraints may be either mitigated or corroborated with more presence data outside of the Lockyer's Waters and Bay d'Espoir regions of the island. The first assumption: if *E. pedicellatum* is found beyond the geographic and biogeographic constraints (i.e. beyond the 20 km hinterland region), then the new data can be used to build a more robust model with a modified distance

predictor. My second assumption: if *E. pedicellatum* is not found beyond the 20 km hinterland boundary in areas that lack the geographic and biogeographic constraints, this may, in turn, strengthen the distance from coastline predictor. In the summer of 2010, new *E. pedicellatum* were found in the Bay du Nord region just northeast of the Bay d'Espoir population, which may possibly reinforce my first assumption.

An additional confounding parameter might be the fact that there are different measurements of distance from coastline. Newfoundland is an irregularly shaped landmass that conforms to the fractal geometric shape of most coastlines. Instead of using straight distance from coastline, a better metric would be to develop a fractal scale that indicated exposure to coastline, rather than distance from coastline. This may eliminate straight line distance as a factor and be more sensitive to the amount and exposure of the indented Newfoundland coastline for any given presence or absence data point. This may also help alleviate the issue of geographical factors of the Avalon data set, as the Avalon would have the highest overall exposure rate to coastline than any other part of the island, yet not be dependent upon straight line distance. I had originally tried to develop a fractal weighting scale for my model, but it was not practical for interpretation in ESRI ArcGIS 9.3 at this time. Better programming skills or a dedicated AML program or process in ArcGIS may rectify this issue.

The measurement of more effective moisture data that directly pertains to the life history of *E. pedicellatum* is strongly needed. Small, inexpensive weather data gathering stations could be placed in the hyper dense populated regions of Bay d'Espoir and Lockyer's Waters to monitor humidity, as well as fog and rainfall. Bill Clarke (pers. comm.) has suggested that the months of July to September are the critical time to

measure moisture levels, as these months would be most likely to be limiting in terms of air and aerosol moisture content because they are the driest months in the year.

The use of another predictor based on newer evidence of population patterns and life history could also be used in further studies. Claudia Hanel (pers.comm.) has suggested that proximity to bogs and fens may be a key component. Hanel's expert opinion is that the growth cycle of the balsam fir near the edges of fens and bogs is severely stunted (due to nutrient-poor, or acidic soils), and it may affect the period of time that the bark of the balsam fir retains its smooth features. Because it is believed that *E. pedicellatum* requires approximately 30 years to go through an entire life cycle, the rate at which balsam fir matures could be a factor. Trees that are growing very quickly and progressing through the lifecycle (and hence reducing the bark substrate's optimal thallus attachment time period) may have less overall ability to provide suitable substrate for *E. pedicellatum*. Therefore, a study on tree age, DBH (diameter at breast height) and growth patterns could be used to determine if there is significant relationship between rate of substrate change to *E. pedicellatum* population density.

The use of other possible indicator species in proximity to *E. pedicellatum* may be another way to strengthen the model. The Nova Scotia heuristic model, developed in 2008, used a suite of other lichen species in determining the possibility of detecting *E. pedicellatum* (Cameron and Neily, 2008). In Newfoundland, other lichen and fungal species such as *Cococarpia spp.* have been detected near *E. pedicellatum* populations. It has also been suggested by some researchers that a possible relationship exists between *E. pedicellatum* and certain liverwort species *(Frullania* spp.) during part of *E. pedicellatum*'s lifecycle (Michelle Piercey-Normore, pers. comm). In particular, the

water lobes of the liverwort may contain the photbiont of *E. pedicellatum*, and may play a role in the union between fungi and algal bionts (Michelle Piercey-Normore, pers. comm). McCarthy's field research in 2008 was an example of a multi-species lichen survey in which a suite of lichen species data was gathered for any given survey area (McCarthy, pers. comm.). Similar surveys in the future that focused on a suite of lichens may be able to determine common lichen guild associations with *E. pedicellatum*. Except for Dr. McCarthy's field survey, most of the information gathered about other lichen and fungi species in close proximity to *E. pedicellatum* presence data has been opportunistic and sporadic in nature. This is due to the focus of surveys being to determine *E. pedicellatum* population and to the amount of training required for identification of a large suite of potentially associated lichen species. A few species, such as *Cococarpia* spp., that are readily identified and known to occur in similar habitats have occasionally been included in surveys.

Using systematic survey practices rather than the current opportunistic search practices would be useful in the future. Systematic survey efforts would look in regions outside of the hyper-dense populated regions found so far. As stated in my introduction, the real-life feasibility of such practices is not trivial, as access to remote regions, transportation, weather, and logistical concerns would limit the financial and time related value of a large scale field survey of this nature. Dr. McCarthy's systematic survey of lichens on the Avalon in 2008 was the first such attempt to do this on a large scale, but only lasted for one year (McCarthy, pers. comm.). The Avalon itself is easier to survey in such a fashion, as roads and access to remote areas is more readily available, than the rest of the island. The survey in the Bay du Nord region (summer of 2010) typifies the

opportunistic surveying practice for both practicality and time efficiency, as regions deemed suitable for *E. pedicellatum* were chosen rather than a systematic search pattern due to the cost and remoteness of such a survey endeavour.

In future surveys, the exclusion of site counts in favour of point counts would add more fine-resolution data to the data set and hopefully increase statistical fitness of the final model. If site counts are continued, a notation of survey time spent, number of surveyors, and/or approximate area searched may be an asset in including such data points into the final statistical model for both training and testing purposes. However, if the grain size that I had used for my study (78.9 meters) was smaller for a newer study, this would further compound the usage of site data, as surveyor accuracy and precision of their estimation of time spent and area surveyed would come into question.

The need to do more absence data recording would further help future model development. Pseudo absence data are only a proxy for true absence data, and as such are not as statistically robust as true absence data. The inclusion of absence data both in government lichen surveys, and in research would hopefully decrease reliance on pseudo absence data and increase the fitness of the overall model.

On a final note, further modeling should be coupled with field research. The advantage of my research was the wealth and quantity of data that had been accrued by experts and researchers for more than 10 years. However, as mentioned in my above recommendations and limitations of my research, key elements such as absence data, moisture data and systematic data practices were missing. Furthermore, some of the recommendations I have made (e.g., tree age comparisons) would be better facilitated by hands-on field data from the researcher, rather than relying on data that have been

accrued by other researchers, but lacked the detail need for that particular focus for their study. I did not have a field research component in my study, so I had to rely on 3rd party data and proxies (even though the detail, quality and quantity was extensive). Therefore, budgeting for field work and research, as well as equipment and transportation and lodging should be considered for future research projects.

5 Conclusion

The population distribution of *Erioderma pedicellatum* on the island of Newfoundland can be mapped using a systematic predictive model. Over 10 years of data collected by researchers and surveyors has given a clear picture of the population distribution in the Bay d'Espoir and Lockyer's Waters population regions. These data, in turn, were used to develop a model based on three final predictors: distance from coastline, topographical aspect and relative near surface moisture.

Due to the non parametric nature of the population distribution of *E. pedicellatum*, the distance from coastline predictor showed a clear bimodal distribution. The use of the Generalized Additive Model was used to statistically analyze all three predictors. Eighteen candidate models were developed based on various combinations of these three predictors and a final model using distance from coastline and topographical aspect was chosen. The model proved relatively poor in determining the presence of *E. pedicellatum*, but was relatively accurate at predicting absence regions. The confusion matrix analysis gives the model poor standings in overall model fitness.

While all three final predictors were deemed individually significant in the model building phase of the research, a final review concludes that these predictors should be

refined in the future with more detailed in-situ data (for moisture) or a more accurate metric (exposure to coastline versus distance) and possibly another unknown model parameter which may be an important factor for increasing model fitness and sensitivity. Developing a predictive model based on little life history information for a species that is hard to find and rare has been a challenge. This was a preliminary model using the data at hand and determining predictors based on expert opinions, and by gleaning the current data sets. The recent *E. pedicellatum* surveys outside of the two hyper dense populated regions (Bay du Nord), coupled with more life-history knowledge of the Newfoundland *E. pedicellatum* populations may add better quality data to a more refined future model.

References

Ahti, T. 1983. Lichens. Pages 319–360 *in* G. R. South, editor. Biodiversity and Ecology of the Island of Newfoundland. Junk Publishers, The Hague.

Anderson, D. R., (2008). *Model Based Inference in the Life Sciences*. Springer Science + Business Media LLC. New York.

Ayers, T. 2010. Rare, endangered lichens found on super tree in Richmond County. The Cape Breton Post, January 14.

Baldwin L. K., Bradfield G.E., (2005). Bryophyte community differences between edge and interior environments in temperate rain-forest fragments of coastal British Columbia. *Canadian Journal of Forestry Resources, 35*, 580-592.

Beyer, H. L. 2004. Hawth's Analysis Tools for ArcGIS. Available at http://www.spatialecology.com/htools.

Bratko, A. Filipic, B. Cormack, G., Lynam, T., Zupan, B., (2006). Spam filtering using statistical data compression models. *The Journal of Machine Learning Research*, 7, 2673-2698.

Brodo, I.M., Sharnoff, S.D., Sharnoff, S., (2001). *Lichens of North America*. New York, USA, Yale University Press.

Cameron, R.P., Neily, T., (2008). Heuristic model for identifying the habitats of *Erioderma pedicellatum* and other rare cyanlichens in Nova Scotia, Canada. *The Bryologist*, 111 (4), 650-658.

Campbell, J., Coxson, D. (2001). Canopy microclimate and arboreal lichen loading in subalpine spruce–fir forest. *Canadian Journal of Botany*, 79, 537–555

COSEWIC, 2002. *Lichen, Boreal Felt*. Retrieved from the web http://www.cosewic.gc.ca/eng/sct1/searchdetail_e.cfm.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC., Peterson, A.T., Philips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmerman, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129-151.

Engler, R., Guisan, A., Rechsteiner, L., (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, 41, 263-274.

Fielding, A.H., Bell, J.F., (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24 (1), 38-49.

Franklin, J., (1995). Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 1995, 19, 474-499.

Gauslaa, y., Ohlson, M., Solhaug, K.A., Bilger, W., Nybakken, L., (2001). Aspect-dependent high-irradiance damage in two transplanted foliose forest lichens,

Lobaria pulmonaria and *Parmelia sulcata. Can. Journal of Forest Research,* 31, 1639–1649.

Gillis, M., (2001), Canada's national forest inventory (responding to current information needs). *Environment Monitoring and Assessment*, 67, 121-129.

Gu, W., Swihart, R.K., (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, 116, 195-203.

Guisan, A., Edwards, T., Hastie, T., (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling,* 157, 89-100.

Guisan, A., Thuiller, W., (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993-1009.

Guisan, A., Zimmerman, N.E., (2000). Predictive habitat distribution models in ecology. *Ecological Modeling,* 135, 147-186.

Hastie, T., (2010). Package 'gam': *Generalized Additive Models*. Retrieved from the web from http://cran.r-project.org/web/packages/gam/gam.pdf .

Hastie, T., Tibshirani, R., (1986). Generalized additive models. *Statistical Science*, 1, 297-318.

Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N., (2002). Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, 83, 2027-2036.

Hylander, C. (2005). Aspect modifies the magnitude of edge effect on bryophyte growth in boreal forests. *Journal of Applied Ecology,* 42, 518-525.

Kaplan, R.S., Urwitz, G. (1979). Statistical models of bond ratings: a methodological inquiry. *Journal of Business*, 52, 231-261.

Keeping, B., Hanel, C., (2006). *A five year (2005-2011) management plan for the boreal felt lichen (Erioderma pedicellatum) in Newfoundland and Labrador.* Government of Newfoundland and Labrador.

Lang, W., Reiners, A., Pike, L., H., (1980). Structure and biomass dynamics of epiphytic lichen communities of balsam fir forests in New Hampshire. *Ecology*, 61, 541-550.

Maass, W.S.G and Yetman, D. 2002. COSEWIC Assessment and Status Report on the Boreal Felt Lichen, *Erioderma pedicellatum,* in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa.

Meades, W.J., Moores, L., (1989). *Forest Site Classification Manual. A Field Guide to the Damman Forest Types of Newfoundland.* 1st Edition. Newfoundland Department of Forestry and Agriculture, Forestry Canada.

Moores, L., Pittman, B., Kitchen, G., (1996). Forest ecological classification and mapping: their application in ecosystem management in Newfoundland. *Environmental Monitoring and Assessment*, 39, 571-577.

Newfoundland and Labrador, Department of Environment and Conservation (2010). *Species at Risk*. Retrieved from the web at http://www.env.gov.nl.ca/env/wildlife/endangeredspecies/index.html on May 29th, 2010.

Parker, A.J., (1982). The topographic relative moisture index: An approach to soil-moisture assessment in mountain terrain. *Physical Geography*, 3, 160-168.

Pearson, R.G., Raxworhty, C.J., Nakamura, M., Peterson, A.T., (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography,* 34, 102-117.

Rolstad, J., Gjerde, I., Storaunet, K.O., Rolstad, E., (2001). Epiphytic lichens in Norwegian coastal spruce forests: historic logging and present forest structure. *Ecological Applications*, 11, 421-436.

Sattler, T., Bontadina, F., Hirzel, A., Arletta, R., (2007). Ecological niche *modeling* of two cryptic bat species calls for a reassessment of their conservation status. *Journal of Applied Ecology*, 44, 1188-1199.

Seaward, M.R.D., (1977). *Lichen Ecology.* New York, USA: Academic Press.

Scheidegger, C., (2003). *Erioderma pedicellatum*. In: IUCN 2008. 2008 IUCN Red List of Threatened Species. Retrieved November 22nd, 2008 from www.iucnredlist.org.

Urban, D.L., (2000). Using model analysis to design monitoring programs for landscape management and impact assessment. *Ecological Applications,* 10, 1820-1832.

Werth, S., Tommervik, H., Elvebakk, A., (2005). Epiphetic macrolichen communities along regional gradients in northern Norway. *Journal of Vegetation Science*, 16, 199-208.

Wilds, S., van Manen, F., (1995). Arc Markup Language script for the Topographic Convergence Inventory. Created by S. Wilds, and modified by F. van Manen. USGS.

Wisz, M., Guisan, A., (2009). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecology*, 9. Retrieved from http://www.biomedcentral.com/1472-6785/9/8.

Wolock, D.M., G.J McCabe, (1995). Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research,* 31, 1315-1324.

Zaniewski, A.E., Lehmann, A., Overton, J., (2002). Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modeling*, 157, 261-280.